# Somatic Hypermutation Introduces Insertions and Deletions into Immunoglobulin V Genes

By Patrick C. Wilson,*‡ Odette de Bouteiller,§ Yong-Jun Liu,§ Kathleen Potter,* Jacques Banchereau,‖ J. Donald Capra,*‡ and Virginia Pascual*

*From the *Molecular Immunology Center, Department of Microbiology, University of Texas Southwestern Medical Center at Dallas, Texas 75235-9140; the ‡Oklahoma Medical Research Foundation, Oklahoma City, Oklahoma 73104; §Schering-Plough Laboratory for Immunological Research, Dardilly, France; and ‖Baylor Institute for Immunological Research, Dallas, Texas 75235*

## Summary

During a germinal center reaction, random mutations are introduced into immunoglobulin V genes to increase the affinity of antibody molecules and to further diversify the B cell repertoire. Antigen-directed selection of B cell clones that generate high affinity surface Ig results in the affinity maturation of the antibody response. The mutations of Ig genes are typically base-pair substitutions, although DNA insertions and deletions have been reported to occur at a low frequency. In this study, we describe five insertion and four deletion events in otherwise somatically mutated $V_H$ gene cDNA molecules. Two of these insertions and all four deletions were obtained through the sequencing of 395 cDNA clones (~110,000 nucleotides) from $CD38^+IgD^-$ germinal center, and $CD38^-IgD^-$ memory B cell populations from a single human tonsil. No germline genes that could have encoded these six cDNA clones were found after an extensive characterization of the genomic $V_H4$ repertoire of the tonsil donor. These six insertions or deletions and three additional insertion events isolated from other sources occurred as triplets or multiples thereof, leaving the transcripts in frame. Additionally, 8 of 9 of these events occurred in the CDR1 or CDR2, following a pattern consistent with selection, and making it unlikely that these events were artifacts of the experimental system. The lack of similar instances in unmutated $IgD^+CD38^-$ follicular mantle cDNA clones statistically associates these events to the somatic hypermutation process ($P = 0.014$). Close scrutiny of the 9 insertion/deletion events reported here, and of 25 additional insertions or deletions collected from the literature, suggest that secondary structural elements in the DNA sequences capable of producing loop intermediates may be a prerequisite in most instances. Furthermore, these events most frequently involve sequence motifs resembling known intrinsic hotspots of somatic hypermutation. These insertion/deletion events are consistent with models of somatic hypermutation involving an unstable polymerase enzyme complex lacking proofreading capabilities, and suggest a downregulation or alteration of DNA repair at the V locus during the hypermutation process.

During the course of a T cell-dependent antibody response, B cells hone the specificity of their antibody molecules through a process of random somatic hypermutation of their V genes, followed by antigen driven selection. This is collectively referred to as affinity maturation. This process occurs within the germinal centers (GCs)[1] of secondary follicles from peripheral lymphoid organs when antigen stimulated B cells receive proper signals from T and accessory cells. In the human system, GC B cells are characterized by the surface expression of CD38 and, in most cases, the loss of IgD (1–3). We have previously shown that the initiation of somatic hypermutation occurs within the $CD77^+$ subset of these $IgD^-CD38^+$ B cells (4). Mutated V genes can be isolated from all subsequent stages of B cell differentiation and in cells from all $IgD^-$ and certain $IgD^+$ B cell subsets (4, 5). The molecular process of somatic hypermutation remains elusive, primarily due to the lack of a good in vitro model until very recently (6). Much of what

[1] *Abbreviations used in this paper*: FM, follicular mantle; FW, framework; GC, germinal center.



J. Exp. Med. © The Rockefeller University Press • 0022-1007/98/01/59/12 $2.00
Volume 187, Number 1, January 5, 1998 59–70
http://www.jem.org



***EXHIBIT C***

is known concerns: (*a*) localizing the somatic hypermutation process to particular B cell subsets and anatomical settings (4, 7–10); (*b*) delineating the limits and rates of mutational activity (11); (*c*) determining the minimal substrate through transgenic technology (12, 13); and (*d*) analyzing the mutations themselves in the context of the surrounding sequence to reveal tendencies such as strand polarity and "hotspots" of somatic hypermutation (for reviews see references 12 and 13).

Although somatic hypermutation is typically described as the generation of bp substitutions, insertions and deletions have been sporadically described. As with somatic point mutations, the analysis of these events can provide valuable information concerning somatic hypermutation itself. Analysis of human $V_H4$ family genes generated from the amplification of cDNA from somatically mutated GC ($IgD^-CD38^+$) and memory ($IgD^-CD38^-$) B cell subpopulations led us to identify a number of cDNA clones from the mutated cell populations that contained insertions and deletions. We provide evidence that these events are linked to the somatic hypermutation process. Additionally, these events occur in a predictable fashion relative to the surrounding sequence, suggesting a model for their occurrence with implications for the molecular process of somatic hypermutation.

## Materials and Methods

*Isolation, Labeling, and Sorting of Tonsil B Cells.* Human tonsils were obtained during routine tonsillectomy. B cell isolation and sorting for CD38 and IgD expression were performed as previously described (4, 14). In brief, human tonsillar B cells were separated into $IgD^+CD38^-$ follicular mantle (FM) B cells, $IgD^-CD38^+$ GC B cells, and $IgD^-CD38^-$ memory B cells to 95–98% purity as predicted by FACS® analysis, as previously described (13). The mutation state of the $V_H$ gene cDNA clones from the various subpopulations was in agreement with our previous study (4). Clones were considered somatically mutated if they contained two or more bp substitutions, well beyond the expected error rates for the avian myeloblastosis virus reverse transcriptase (AMV-RT), Taq, and PFU polymerases used in these analyses (this mutation rate is based on our previous analyses; reference 4).

*Sequencing the Ig $V_H$ Transcripts.* Total RNA was extracted from $1–5 \times 10^5$ B cells using guanidinium thiocyanate-phenol-chloroform in a single step using the Ultraspec RNA isolation system (BIOTECX Laboratories, Houston, TX), and was reverse transcribed using oligo-d(T) or specific V gene constant region oligonucleotides Cμ12 (5′-CTGGACTTTGCACACCAC-GTG-3′) for IgM transcripts or Cγ180 (5′-CTGCTGAGG-GAGTAGAGTCC-3) for IgG transcripts and SuperScript II reverse transcriptase (GIBCO BRL, Gaithersburg, MD). First strand cDNA was used directly for second strand synthesis and amplification via PCR using internal primers corresponding to the Cμ or Cγ constant regions in combination with $V_H4$ or $V_H6$ family-specific leader oligonucleotides: Cγ140, 5′-GGCAAGGTGTGCACGCC-GCTG-3′; Cμ10, 5′-TCTGTGCC CTGCATGACGTC-3′; L-4, 5′-ATGAAACACCTGTGGTTCTT-3′; L-6, 5′-ATGTCTGT-CTCCTTCCTCAT-3′. The PCR products were purified using microconcentrators (Amicon, Beverly, MA), and then were kinased and blunt-end ligated into an EcoRV-digested and dephosphorylated pBluescript plasmid (Stratagene, La Jolla, CA; Polynucleotide Kinase, T4 DNA Ligase, and EcoRV were from Boehringer Mannheim, Amsterdam, Netherlands). After transformation by electroporation into electro-competent DH10α *Escherichia coli* (GIBCO BRL) and screening with consensus internal oligonucleotides as previously described (4, 15), positive colonies were picked, plasmid mini-preparations were made, and colonies were sequenced in both directions using an automated DNA sequencer and automated sequencer protocol (ABI-377; Advanced Biotechnologies Inc., Columbia, MD). All sequences were analyzed using DNAstar (DNAstar Inc., Madison, WI). In the first tonsil analyzed, 583 clones were picked, plasmid mini-preparations were made, and Southern blots were prepared by standard methods. These blots were screened with a set of oligonucleotides specific for the various $V_H4$ family genes. Only those clones that screened positive with constant region probes but negative for the various $V_H4$ complementarity-determining region (CDR)1–specific probes were sequenced (395 of 583 clones), thus enriching the somatically mutated populations analyzed, in that the CDR1 probes should anneal only to the sequences most similar to germline. The frequency of the occurrence of these events can therefore only be predicted to be between 6 out of 395 and 6 out of 583 clones (1–2%). Any sequence of interest was resequenced in both directions to ensure sequence fidelity.

*Characterizing the Genomic Repertoire.* Total genomic DNA was isolated from FM B cells ($IgD^+$, $CD38^-$) using the Puregene DNA isolation kit (Gentra Systems, Inc., Minneapolis, MN). $V_H4$ genes were amplified using a $V_H4$ leader-specific primer (L-4, as above) and a primer specific for all $V_H4$ gene family heptamer–nonamer spacer regions as previously described (16). PCR products were agarose gel purified, then cloned into *E. coli* as described above for the cDNA clones. Clones identified in the cDNA analysis that contained insertions or deletions were used to design PCR primers to amplify both the exact sequence of clones with insertions/deletions as found and the predicted sequences based on the proposed germline counterparts. Oligonucleotides used in this analysis (Format, is as follows: clone: exact/predicted): g64:5′-GGACGG-GTTGTACTTGGTTCC-3′/5′-GGACGGGTTGTAGGTC-TCC-3′; g144:5′-TCTTGAGGGACGGGTTGGTGT-3′/5′-TCTTGAGGACGGGTTGT-3′; g187:5′-CAGCTCCAGTAG-TAAGCCCCG-3′/5′-CAGCTCCAGTAGTAACCACCG; g188: 5′-GAGGGATTGTAGTTGGAGCC-3′/5′-GAGGGGTTGT-AGTTGGTCCC; g192:5′-CCAGCCCCAGTAGTAGTAACT-3′/(same); and g80:5′-GCGGATCCAATACCTCACACT-3′/5′-GCGGATCCAGTAGTAACC-3′.

*Sequence Availability.* All cDNA sequences with insertions or deletions, and any genomic sequences unique to the literature as described in the results section are available from EMBL/Genbank/DDBJ under accession numbers AF013615 through AF013626.

*Assay for Screening $V_H$ Gene Lengths.* To facilitate the analysis of large numbers of $V_H$ gene transcripts for the presence of insertions or deletions, first strand cDNA produced as described above was PCR amplified using Expand high fidelity polymerase (Boehringer Mannheim) to reduce errors resulting from Taq polymerase alone. The products of this PCR amplification were cloned as described above and screened using $^{32}P$-labeled, gene-specific oligonucleotides ($V_H4$-39:5′-ATTGCGAGTATCTATTATAGT-3′; L-6 as above). Positive colonies were picked and used to inoculate overnight cultures. A 1 μl aliquot from each 24-h culture was used to directly inoculate 25-μl PCR amplification mixtures in 96-well-format PCRs. The internal PCR reactions used $^{32}P$-labeled, gene-specific oligonucleotides to amplify a 230-base fragment including the $V_H4$-39 CDR1 (L-4, as above, and $V_H4$-39-3′: 5′-

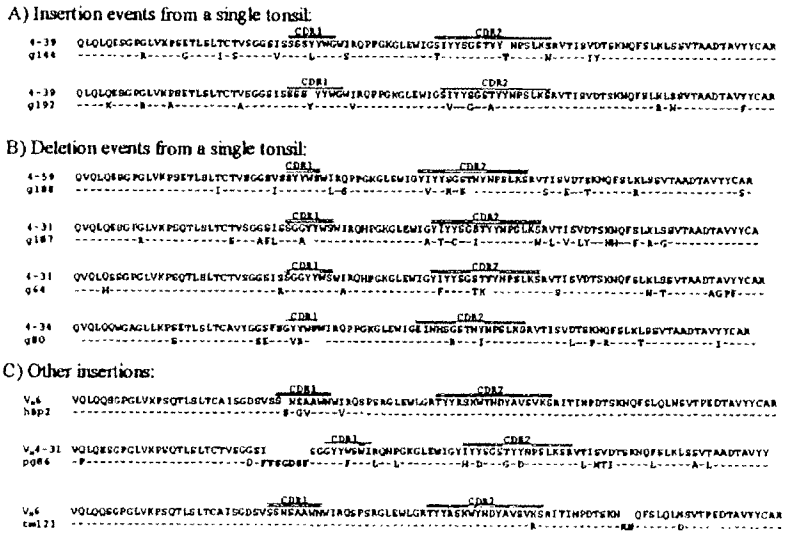


**Figure 1.** Predicted amino acid sequences of nine cDNA clones with insertions or deletions. Note that all of these clones are extensively mutated. In all but clone tm121, the insertions or deletions occur in CDR1 or CDR2. (*A* and *B*) Two clones with insertions (*A*) and four clones with deletions (*B*) from a single tonsil. (*C*) Three additional cDNA clones with insertions isolated from various sources. Sequence data available from GenBank/DDBJ under accession numbers AF013615 through AF013626

GCTCCCACTATAATAGATACT-3′) or for analysis of $V_H6$ genes a 166-nucleotide fragment including the CDR1 and CDR2 of $V_H6$ ($V_H6$FW1: 5′-TGCCATCTCCGGGGACAGTGT-3′, $V_H6$FW3: 5′-TGTCTCTGGGTTCATCGTTAT-3′). Aliquots of each clone were also used to inoculate amplifications of CDR3 regions using FW3-specific (ssFW3: 5′-CTGAA[C/G]CT-GAGCTCTGTGAC[T/C]) and Cμ- or Cγ-specific oligonucleotides (CμD: 5′-GGAATTCTCACAGGAGACGA-3′, Cγ-140 as above) to analyze the diversity of the populations under study; the distribution of CDR3 size variations of several hundred $V_H$ sequences cloned in this analysis were used to produce an expected distribution of CDR3 sizes for comparison (see Fig. 5 *B*). The amplification products were electrophoresed on 0.6X-TBE, 5% urea-acrylamide sequencing gels (Long Ranger; J.T. Baker, Phillipsburg, NJ) and analyzed with a PhosphorImager (Molecular Dynamics, Sunnyvale, CA.) using the Image Quant software supplied by the manufacturer. Clones that differed from the expected size and those clones in lanes adjacent to aberrantly migrated bands were used to produce plasmid preparations from which the inserts were sequenced in either direction.

*Scoring of Insertion/Deletion Events.* In the results section, insertions and deletions are scored as events per $10^4$ nucleotides within the customary boundaries of CDR1 and CDR2. This unit was chosen because in the selected populations studied these events are generally only found in the CDR regions and therefore the comparison of events per total nucleotides would be misleading. In the PAGE analysis, each $V_H4$-39 FM clone included only the CDR1 (21 nucleotides) within a total of 230 nucleotides/clone, whereas each $V_H6$ FM clone was only 166 nucleotides but included both the CDR1 and CDR2 (75 CDR nucleotides). In the sequencing analysis, various B cell populations were analyzed involving a wide range of overall lengths. Comparisons of the frequency of insertions/deletions just within the CDRs allowed for a more standardized and quantitative analysis, and for more freedom in experimental design.

*Baculovirus Expression System.* Cloning and coexpression of clone pg86 and κ light chain FS6κ in the baculovirus expression system was performed as previously described (17). Recombinant *Autographa californica* nuclear polyhedrosis virus (AcMNPV) was cloned using the pH360NX transfer vector and expressed in Sf9 cells.

*Capture ELISA for γ Heavy Chain, and κ Light Chains.* Expression of recombinant antibodies of clone pg86 coexpressed with κ light chain FS6κ were measured by capture ELISA. Wells were coated with goat anti-human IgG and incubated with supernatant of recombinant pg86/FS6κ added in serial twofold dilutions. Bound antibody was detected using alkaline phosphatase-conjugated goat anti-human IgG, or goat anti-human Cκ. After 1-h incubation at 37°C, phosphatase substrate was added and absorbance was measured at 405 nm in an ELISA plate reader.

## Results

*Insertions and Deletions into Immunoglobulin $V_H$ Genes.* In a large scale analysis of $V_H$ genes from both the IgM and IgG compartments of B cell subpopulations separated from a single human tonsil, six clones that contained DNA insertions or deletions were isolated. These insertions and deletions were apparently selected in that they involved nucleotide triplets or multiples of nucleotide triplets, leaving the cDNAs (transcripts) in frame, and they were localized to the CDR1 and CDR2 (Fig. 1, *A* and *B*). The six clones with insertions or deletions were identified from the sequencing of 395 cDNA clones (~110,000 nucleotides) from GC and memory B cell subpopulations, resulting in a frequency of <2% of clones analyzed (~1 event/18,000 nucleotides). All six events were in IgG transcripts. Two events were obtained from $IgD^-CD38^+$ GC and four events from $IgD^-CD38^-$ memory cell populations. None of the IgM $V_H$ cDNAs analyzed from this tonsil had insertions or deletions, although we have observed such events in IgM transcripts in the past and in subsequent analyses, as described below.

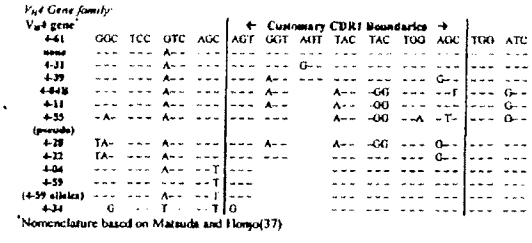


Figure 2. Comparison of the CDR1s of the human $V_H4$ germline genes. The primary variability between $V_H4$ family members is 3–6-bp size variances in the CDR1s which is similar to the short insertions and deletions that we attribute to somatic hypermutation in the selected B cell populations studied in this report.

*The Insertions and Deletions Are Not Germline Encoded.* The analysis described above focused on the $V_H4$ gene family, which consists of 10–14 members/genome, varying slightly between individuals (16, 18). As shown in Fig. 2, the major difference between $V_H4$ genes involves the length of CDR1. Because genomic diversity between $V_H4$ family members resembles the events described in this paper we had to rule out possible alternative explanations for these events, such as: (*a*) different alleles of the detected genes; (*b*) rarely expressed or otherwise unknown $V_H4$ gene family members; or (*c*) hybrids between known and detected $V_H$ genes and/or other artifacts of the experimental system. To address these issues, both the expressed and genomic repertoires from this tonsil were characterized. As indicated in Table 1, 2 out of 118 $V_H4$–39, 2 out of 49 $V_H4$–31, 1 out of 87 $V_H4$–34, and 1 out of 45 $V_H4$–59 cDNA clones contained insertion/deletion events. cDNA clones were judged as unique isolates based on CDR3 analysis, and the few isolates that appeared to be clonally related differed in their patterns of somatic mutation beyond the level explainable by reverse transcription and PCR errors (maximum: >1 mutation/500 nucleotides of $V_H$ gene sequence as previously described [4]).

To characterize the genomic repertoire of the initial tonsil, 80 germline $V_H4$ gene clones were isolated and sequenced (Table 1), which encompassed all 14 $V_H4$ family members or alternate alleles represented in the 446 cDNA clones analyzed from all of the tonsillar B cell subsets. In the course of this study, we isolated the germline counterpart of a novel $V_H4$ gene segment for which transcripts had been found. In addition, germline genes corresponding to two apparently functional $V_H4$ genes not found as cDNA clones in this analysis were isolated, as well as one nonfunctional $V_H4$ gene and a divergent polymorphism of a known $V_H4$ pseudogene. The proposed germline counterparts of each of the $V_H4$ genes containing insertion/deletion events were isolated from 4 to 11 times (Table 1). 8 independent genomic isolates of $V_H4$–31 and of $V_H4$–39 were cloned. $V_H4$–34 and $V_H4$–59 were isolated 11 and 4 times, respectively. No germline genes were isolated that could have encoded the insertion/deletion events described.

To further be certain that the insertion/deletion events described herein were not germline encoded, two sets of PCR primers were designed to specifically recognize: (*a*) the exact sequence of the events; (*b*) the predicted, unmutated, germline sequence corresponding to the cDNAs containing insertion and deletion events. These primers were used to amplify genomic DNA from this individual, yielding negative results (data not shown). The unique nature of these events relative to both the expressed and genomic repertoire and our inability to amplify genomic counterparts for these events by PCR suggest that they are not germline encoded.

**Table 1.** *cDNA and Germline Clones Isolated*

| $V_H4$ gene alleles isolated* | cDNA clones with ins/del | Total cDNA clones isolated | Germline clones isolated‡ |
|---|---|---|---|
| $V_H4$–39 | 2 | 113 | 7 |
| $V_H4$–31 | 2 | 49 | 8 |
| $V_H4$–59 | 1 | 45 | 4 |
| $V_H4$–34 | 1 | 87 | 11 |
| $V_H4$–34 related | 0 | 0 | 4 |
| $V_H4$–55 pseudogene§ | 0 | 0 | 12 |
| $V_H4$–55-related pseudogene§ | 0 | 0 | 3 |
| $V_H4$–04 | 0 | 17 | 7 |
| $V_H4$–04-related pseudogene§ | 0 | 0 | 2 |
| $V_H4$–61 | 0 | 25 | 7 |
| New $V_H4$ gene¶ | 0 | 33 | 3 |
| $V_H4$–04B | 0 | 72 | 1 |
| $V_H4$–28 | 0 | 0 | 1 |

*Nomenclature based on Matsuda and Honjo (37).
‡Nine unusual isolates were also cloned consisting of hybrids of two of the indicated genes, resumably do to PCR artifact. None of these artifacts were altered in size or resembled any of the insertion or deletion events observed.
§Pseudogenes contain stop codons or frameshift mutations and are not expressed.
¶Newly identified $V_H4$ gene is most closely related to $V_H4$–04.

*The Proposed Insertion/Deletion Events Are Not the Result of ($V_H/V_H$) Recombination.* As in most V gene repertoire analyses, we detected hybrid $V_H$ sequences that could be the result of either PCR splicing by overlap extension artifacts, or reciprocal homologous recombination between unrearranged V genes (19). However, none of these likely artifactual events were altered in size such that they resembled the insertion or deletion of DNA described above. A number of artifacts of this type had been isolated in the cDNA analysis as well; such artifacts are common to V gene analyses (20). The cDNA isolates with deletion and insertion events were stringently compared to all germline and cDNA isolates and were found to be unique relative to both the expressed and germline $V_H4$ gene repertoires of this individual, supporting a somatic origin for their occurrence.

*The Insertions and Deletions Are Associated with Somatic Hypermutation.* To determine whether or not these inser-

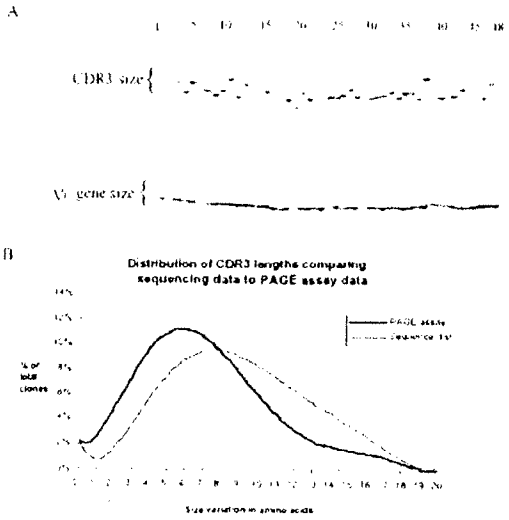


**Figure 3.** Polyacrylamide gel assay to identify insertions or deletions into $V_H$ genes. (*A*) Phosphorimage of a polyacrylamide gel; each lane contains the hot-PCR products ($^{32}$P-labeled) of the $V_H$ gene and the CDR3 of an individual clone. (*B*) A comparison of the distribution of CDR3 sizes of the 485 CDR3s assayed to the distribution of 500 CDR3s observed in sequences from this report indicates that the clones assayed by electrophoresis were a polyclonal population. CDR3 sizes were measured from the most 3' Tyr residue (common to all V genes analyzed) to the most 5' Cμ or Cγ residue. CDR3 lengths for those assayed by electrophoresis were extrapolated based on sequencing of 75 out of the 485 clones assayed. The *x*-axis is the number of amino acids greater than the shortest CDR3 found.

tion/deletion events were associated with somatic hypermutation, we analyzed their occurrence in unmutated FM transcripts. This was done using either direct sequencing or PCR amplification of portions of the $V_H$ genes spanning the CDRs, followed by size comparisons on polyacrylamide gels (Fig. 3). Any clones that ran aberrantly, and the clones in adjacent lanes, were sequenced (75 out of the 485 clones). None of these 75 clones were related based on CDR3 homology. To ensure that the remaining 410 FM clones were polyclonal, the CDR3s were PCR amplified and loaded on the sequencing gels simultaneously to the $V_H$ gene amplification products for size comparisons (Fig. 3 *A*). The size distribution of these CDR3s was similar to that of ~500 $V_H$ gene sequences analyzed in this study (Fig. 3 *B*), providing evidence that our FM sample is polyclonal.

The six events detected from a single tonsil were isolated from 395 mutated cDNA clones (25,482 CDR nucleotides), corresponding to a frequency of 2.35 events/$10^4$ CDR nucleotides. This is significantly different ($p=0.014$ by a one-sided $\chi^2$ test) from the analysis of unmutated FM-derived clones (25,515 CDR nucleotides) that yielded no insertions or deletions (Table 2).

In the course of the analysis described above, we isolated one IgM clone containing a 6-nucleotide insertion into framework (FW)3 (see below). We believe that this clone is part of the mutated GC or memory repertoire because it contained 4 bp substitutions in addition to the insertion. In this study, the B cell populations analyzed were 95–98% pure, and the FM B cell subpopulation could therefore include between 2 and 5% contaminating clones, that is, IgM-expressing cells not from the naive population that can therefore be somatically mutated. However, none of

**Table 2.** *Analysis of Unmutated FM cDNA Clones for Insertion or Deletion Events*

| Clone type | Clones assayed | CDR nucleotides[‡] | Events observed | Frequency[§] | Expected (events/$10^4$ CDR nucleotides)[‖] |
|---|---|---|---|---|---|
| Mutated $V_H4$ clones (GC and and memory B cells) | 395 | 25,482 | 6 | 2.35 events/ $10^4$ nn | |
| Unmutated clones: | | | | | |
| $V_H4$-FM, CDR1* | 265 | 5,565 | 0 | 0 | 1.31 |
| $V_H6$ IgM FM $V_H$ genes* | 220 | 16,500 | 0 | 0 | 3.88 |
| $V_H4$ family FM sequences | 51 | 3,450 | 0 | 0 | 0.81 |
| Total unmutated values | | 25,515 | 0 ($P$ = 0.014)[¶] | 0 | 2.35 events/$10^4$ CDR nucleotides |

*Clones analyzed by hot-PCR/PAGE assay as described in the text.
‡CDR nucleotides are those within the customary bounds of the CDR1 and CDR2. (See Materials and Methods for a more detailed explanation of this unit).
§Events per $10^4$ CDR nucleotides.
‖Expected frequency (events/$10^4$ CDR nucleotides) derived from sequencing data; 6 events in 25,482 CDR nucleotides; 6/(25,482 CDR nucleotides/$10^4$) = 2.35.
¶Statistical analysis: $\chi^2$ test for independence.

A Insertion events:

```
V₆:     GTC TCT AGC --- AAC AGT GCT
hBp2:   GTC TCT AGC AGC AAC AGT GCT

V₄-39:  TAC TAC AAC --- CCG TCC CTC
g144:   TAC TAC AGC AAC CCG TCC CTC

V₄-39:  AGT TAC TAC --- TGG GGC TGG
g192:   AGT TAC TAC TAC TGG GGC TGG

V₆:     TCC AAG AAC --- --- CAG TTC
tm121:  TCC AAG AAC AAG AAC CAG TTC
```

B Deletion events:

```
V₄-31:  GGG AGC ACC TAC TAC AAC CCG
g64:    GGa AcC Aag --- TAC AAC CCG

V₄-31:  TCC ATC AGC AGT GGT GGT TAC
g187:   TtC ATC AGC --- GGG GCT TAC

V₄-59:  TAC AGT GGG AGC ACC AAC TAC
g188:   TAC AGg GGg --- tCC AAC TAC

V₄-34:  TAC TAC TGG AGC TGG ATC CGC
g80:    agg TAc --- --- TGG ATC CGC
```

C Long duplication/insertion:

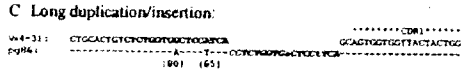

Figure 4. The insertions and deletions are related to the surrounding DNA sequence. (*A*) The insertions involve repetitions of the immediately adjacent sequence. (*B*) The deletions are deletions of tandem repeats. (*C*) The 18-base insertion in clone pg86 is a duplication of the adjacent sequence. Nucleotides that mutated before the duplication/insertion are indicated. Sequence data available from EMBL/GenBank/DDBJ under accession numbers AF013615 through AF013626.

the unmutated FM clones analyzed had insertions or deletions.

*Other Insertions and Deletions into $V_H$ Genes.* We have observed similar instances of insertions and deletions into the coding regions of apparently functional immunoglobulin V genes, including: (*a*) a $V_H6$ IgM isolate containing a triplet duplication/insertion into the CDR1 in addition to several bp substitutions (Figs. 1 *C* and 4 *A*), which was derived from a human hybridoma secreting high affinity mAb against *Bordetella pertussis* (21, 22); (*b*) a 6-nucleotide insertion into the FW3 region of a mutated IgM $V_H6$ gene, representing the only insertion or deletion observed outside of the CDRs (Figs. 1 *C* and 4 *A*, clone tm121); and (*c*) an 18-nucleotide duplication/insertion into a human plasma cell cDNA transcript at the boundary between the FW1 and CDR1 (Figs. 1 *C* and 4 *C*), doubling the length of this hypervariable loop. The viability of clone pg86 was tested by expressing it in the baculovirus system in association with a κ light chain encoding construct (FS-6κ; Fig. 5). The efficient expression, secretion, and pairing with light chain in the baculovirus system suggest that the product of clone pg86 is a functional heavy chain despite the large duplication/insertion.

*The Insertions and Deletions Are Related to the Surrounding Sequence.* As shown in Fig. 4, the insertions reported are duplications of the immediately adjacent sequence, and the

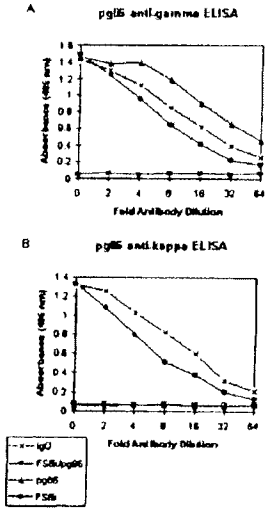


Figure 5. ELISA assays showing the expression of clone pg86 with a six amino acid insertion at the FW1/CDR1 junction. Clone pg86 (IgG heavy chain) was coexpressed with the κ light chain FS6κ in insect cells using the baculovirus expression system. Expression of pg86 and its ability to pair with the FS6κ light chain was tested using capture ELISAs. Wells were coated with goat anti-human IgG. Supernatants containing recombinant antibodies were added in serial twofold dilutions. Bound antibody was detected with phosphatase-conjugated goat anti-human IgG (*A*), and goat anti-human Cκ (*B*).

deletions involve elements of repetitive tracts. In addition, a higher incidence of these events involve sequence motifs that resemble intrinsic hotspots of somatic hypermutation (12, 23–27): (*a*) four of eight events involved the serine codon AGC that has been reported as the "hottest" of hotspots (24–27) (Fig. 4, sequences HBp2, g187, g188, and g86); (*b*) two events involved TAC motifs (Fig. 4, g192 and g64); and (*c*) two events involved the motif AAC (Fig. 1, g144, and tm121). In general all of the clones found to contain insertions and deletions were highly mutated (Fig. 1). Several of these clones had bp substitutions clustered with the insertions or deletions (Figs. 1 and 4). The plasma cell transcript depicted in Fig. 4 *C* contained an 18-nucleotide insertion that duplicated the 5′ adjacent sequence. The central nine nucleotides of the duplicated sequence form a partial palindrome (..GGtGaCtCC..). This clone was mutated (G to A at position 80 and an A to T at position 85) before the duplication/insertion event, as these mutations were perpetuated in the inserted sequence.

## Discussion

Somatic modification of V genes encoding immunoglobulin and T cell receptors recapitulates most mechanisms observed in the evolutionary diversification of DNA: (*a*) V gene recombination, including imprecise junctions, P nucleotides, and untemplated N nucleotide addition; (*b*) gene conversion; and (*c*) bp substitutions in Ig somatic hypermutation. The insertion and deletion of nucleotides is another means for the evolutionary diversification of DNA, and has been proposed as an explanation for unusual V gene sequences in the past (Table 3). In this study, we show that insertions and deletions are associated with the somatic hypermutation process.

*Complexities of the Analysis of Insertions and Deletions into V Genes.* The formal characterization of these events has been a daunting task because of their low frequency, and the complexity of the germline $V_H$ repertoire. According to our study, these events occur in <2% of somatically mutated clones. As shown in Fig. 2, the primary variability between $V_H4$ family members is 3–6-bp size variances in the CDR1s, which is comparable to the short insertions and deletions that we attribute to somatic hypermutation (in selected B cell populations). The similarity between evolutionary diversity and somatic diversification was expected, as the molecules are likely subject to the same functional and structural constraints. This has made it difficult to determine whether these events were generated somatically, versus germline encoded, or if they were artifacts of the experimental system: they could result from homologous recombination between alternate alleles or imperfect recombination between identical alleles, or they could have occurred during B cell replication independent of somatic hypermutation. In fact, $V_H$ genes may exhibit particularly unstable sequence characteristics evolved to help support both germline diversity and the generation of somatic mutations, as suggested by the identification of intrinsic hotspots of somatic hypermutation within the CDRs of V genes (25, 26). Perhaps the area of greatest contention in this complex system remains the possibility that these low frequency events are artifacts of the experimental manipulations performed, the AMV-RT, Taq, or PFU polymerases, and/or the cloning in *E. coli*.

*The Insertion/Deletion Events Are The Result of the Somatic Hypermutation Process.* Our system addresses several key issues that associate the occurrence of insertions and deletions to the somatic hypermutation process. (*a*) Six of the nine insertions/deletions were identified within the $V_H4$ gene repertoire of a single tonsil, providing an experimental system that could be characterized extensively as described below. (*b*) All of the insertion/deletion events reported involved triplets or multiples of triplets, leaving the transcripts in frame and therefore functional, and eight of nine events reported were localized to the CDRs. As with somatic point mutations, no insertions or deletions were observed in the 80 to 120 nucleotides of constant region ($C\mu$ or $C\gamma$) DNA sequenced with each cDNA clone. These hallmarks of somatic hypermutation and selection argue strongly that these events are not artifacts. (*c*) The B cells analyzed were processed and separated into highly pure, mutated B cell populations including GC ($IgD^-CD38^+$) and memory ($IgD^-CD38^-$) B cells, and an unmutated FM B cell population ($IgD^+CD38^-$), making it possible to focus our analysis on the mutated populations and use the unmutated population as a negative control, which in turn allows the statistical association of the observed insertion and deletions to the somatic hypermutation process ($P = 0.014$). In addition, the isolation of four of the insertion/deletion events from memory B cells provides evidence that these events did not result from artifacts related to contamination from endonucleolytically cleaved DNA from the apoptotic GC cells. (*d*) Seven of nine events reported in this study involved $\gamma$ heavy chains that contain nearly twice the mutations of $\mu$ heavy chains (4), further correlating the events described here to somatic hypermutation. (*e*) As discussed below, the insertion/deletion events described tended to involve sequence motifs resembling previously described hotspots of somatic hypermutation, providing evidence that these events occur by the same process. (*f*) Finally, we extensively analyzed the $V_H4$ gene family of the tonsil donor at both the expressed and genomic levels, facilitating the assignment of the insertions/deletions as somatic rather than germline encoded. 6 of the clones with insertions and deletions were unique among 395 $V_H4$ cDNA clones sequenced from a single tonsil, including many independent isolates of each of the $V_H4$ genes expressed (Table 1). In addition, we were unable to isolate genomic templates for any of the insertion or deletion events either by PCR or through the extensive characterization of the genomic $V_H4$ repertoire of the tonsil donor (Table 1). Templating of these events from any other $V_H$ gene family can also be ruled out as members of the seven human $V_H$ gene families differ significantly in the CDR sequences where the events described had occurred.

*Structural and Functional Considerations of Insertions and Deletions into $V_H$ Genes.* The events involving the insertion or deletion of a single amino acid from the CDR1 or CDR2 would not be expected to profoundly alter the backbone structure of these molecules, as the CDRs are the most malleable portions of antibodies. The clone g80 has two of the five amino acids that are customarily considered its CDR1 deleted, leaving only three amino acids to form this hypervariable loop (Fig. 1 *B*). Thus, this is one of the shortest CDR1s reported to date. The clone tm121 has two amino acids inserted into the FW3 region. The portion of the FW3 where this insertion occurred is believed to be solvent exposed and corresponds to the region where the B cell superantigen staphylococcal protein A binds to most $V_H3$-encoded Ig molecules (28); therefore, it is likely that the insertion into this $V_H6$ clone can be tolerated as a loop or bulge on the molecule's surface. The most complex structural change observed in our study involved clone pg86, with a six amino acid insertion at the FW1/CDR1 junction that would presumably double the length of this hypervariable loop and require dramatic structural accommodation. However, we were able to express this heavy chain and found it paired with light chain, indicating that it is likely functional (Fig. 5). The clone HBp2, containing a triplet insert into its CDR1, is particularly interesting because it has a known specificity. This $V_H6$ gene was isolated from a human B cell hybridoma with anti-*Bordetella pertussis* specificity (21, 22). Clone HBp2 has also been expressed in the baculovirus system and is fully functional. We are currently performing mutational analysis of this heavy chain molecule to determine if the additional inserted amino acid plays a role in the affinity and/or specificity of this antibody.

*Analysis of Insertions and Deletions Reported in the Literature.* Various groups have reported a number of insertion and deletion events (Table 3). Virtually all of the insertions

*Table 3. Insertions and Deletions into Somatically mutated V Genes Reported in the Literature*

| Name | Source | Ins/Del (position) | Relation to surrounding sequence | References |
|---|---|---|---|---|
| *Selected populations or coding regions:* | | | | |
| L4-1e | Human $V_H$4-34 (4.21) | ACC insert (within CDR2) | 4-34: AGC ACC AAC (RT)<br>L4-1e: AGg ACC ACC AAC (RT) | 38 |
| 3B62 | Murine $V_H$ 186.2 | GTT deletion (CDR2) | VH186.2: AGT GGT GGT ACT (RT)<br>3B62. AGT GGT ACT | 39 |
| *Unselected populations or untranslated regions:* | | | | |
| 3B62 | Murine $V_H$ 186.2/D/JH2 to JH4 | ACT deletion (3′ untranslated) | GL: GTG ACT ACT TTG (RT)<br>3B62: GTG ACT TTG | 40 |
| 3B44 | Murine $V_H$ 186.2 | 4 single-base deletions (leader intron) | VH186.2: GGC. GGT (RT)<br>3B62: G C. G T | 39, 40 |
| | | | Other 2 events unrelated | |
| M167 | Murine VH107/DFL16.1/JH1 | 2 single-base insertions (leader intron) | GL: ATAG AAGATTAGTAG (RT)<br>M167 ATAGTAAGATTAGTAG | 41 |
| | | (3′ untranslated) | GL: TTTC AGGTCATGAAGGA (RT)<br>M167: TTTgAAGGTCATGAAGGA | 29, 41* |
| | | 5 single-base deletions | GL: GCTTTTC TGTA...CCCAGGAAAAGA<br>M167 GCTTTTC TGTA...CCCAG AAAAGA (IR) | |
| | | (all in 3′ untranslated region) | GL: CTTTTTTCTT (RT)<br>M167 CTTTT TCTT | 41 |
| | | | GL: AGATTTTAC (RT)<br>M167 AGAccT AC | |
| | | | GL: TCATTGG (RT)<br>M167 TCAT GG | |
| | | | GL: GTGACTACTTTGACTACTG (RT)<br>M167 GT ACTACTTTGACTACTG | |

Continued

**Table 3.** *Continued*

| Name | Source | Ins/Del (position) | Relation to surrounding sequence | Reference |
|---|---|---|---|---|
| | | TA deletion (3') | None found (possible hotspot) | |
| | | GTGT deletion (leader intron) | GL: TCTGTGTGTGTGTAT (RT) | |
| | | | M167: TCTGTGT GTAT | |
| M603 | Murine $V_H$S10 7/DFL16.1/JH1 | TC Deletion (leader intron) | GL: TTTTCCTTGTCCTTGTTTT (RT) | 42, 41 |
| | | | M603: TTTTCCTTG CTTGTTTT | |
| | | AAAT deletion (3' Ei/MAR) | GL: GCATTTCTAAAATAAGTTGAGGA (IR) | |
| | | | M603 GCATTTCTA AAGTTGAGGA | |
| MC101 | Murine $V_H$Q52/D/$J_H$3 | GG deletion (3' untranslated) | GL: AAACGGGAATC (RT) | 42, 43 |
| | | | MC101: AAAC GAATC | |
| M511 | Murine Vκ167/Jκ5 | GAA deletion (3' untranslated) | GL: TTTGAAGATAAA (RT) | 42, 44 |
| | | | m511: TTT GATAAA | |
| H37-65 | Murine Vκ Vκ21E/Jκ1-Jκ2 | 11 base deletion (Jκ1/Jκ2 intron) | GL: AGGGACACCAGTGTGTGTACAC (IL) | 45 |
| | | | H37: AGGGc GTACAC | |
| 296.4C11, 253.12D3 | Murine JκC intron | 7 base deletion and a 154 base deletion | No good relationship to surrounding sequence | 46 |
| 2G7 | Murine transgene | single base deletion, and a 49 nucleotide deletion | CTTTGAAGAT . . (N30) . CAGATCAAAG (Repeats form ends of deleted "loop") (IL) | 47 |
| 85κ | Human myeloma Vκ genes | single-base (T) insertions into the CDR1/FW2 junction rendering genes out-of-frame | No relation, however, event followed the proposed hotspot motif TAC | 48 |
| HF-1[†] clone A6: several | Human lymphoma ($J_H$ untranslated) | AG insertion into $V_H$3' untranslated region | Consensus[‡]: GGGGCAG GGC (RT) | 49 |
| | | | clone A6: GGGGCAGAGGGC | |
| | | 30 base deletion | No association | |

RT, repetitive tract; IR, inverted repeat (loop with local DNA); IL, internal loop. *Secondary structure reported by Golding et al. (29). [‡]This study is difficult to interpret in the context of the current report as the genomic $J_H$ locus was not available. The 10 clones were only 80% homologous to the closest $J_H$ locus reported in the literature with most alterations being similar between all of the isolates. Therefore, only 2 of the 26 proposed insertions/deletions can be attributed to somatic mutation with certainty, as they were unique to the consensus of the individual clones.

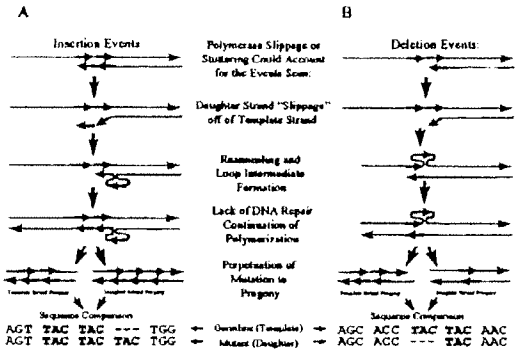


**Figure 6.** Proposed mechanism causing insertion/deletion events: polymerase slippage. This Figure is based on model a of Streisinger et al. (30) and Ripley (31). The same model can account for both (*A*) insertions and (*B*) deletions.

and deletions reported from somatically mutated V genes involved the untranslated regions or occurred in silent passenger transgenes. 19 out of 25 insertions or deletions into somatically mutated genes involved predominantly repetitive elements, or in several cases other sequence patterns associated with secondary structures such as internal homologies or inverted repeats (Table 3). With the inclusion of the 9 events described in this work, 28 out of 34 insertions and deletions involved such elements. Thus, the proximity of sequence elements that can be predicted to cause secondary structural changes in the DNA seems to be a hallmark of insertions and deletions into somatically mutated $V_H$ genes.

*A Model for the Occurrence of Insertions and Deletions during Somatic Hypermutation.* The evidence for the involvement of DNA secondary structure in the production of insertion or deletion mutations during somatic hypermutation, as suggested in 1986 by Golding et al. (29), now seems unequivocal. The insertions and deletions described in our study, and those illustrated in Table 3, occur in a predictable fashion, involving sequence motifs that could form loop intermediates reminiscent of the replication slippage model of Streisinger et al. (30) and Ripley and Glickman (for review see 31) as presented in Fig. 6. Such mutations are postulated to occur when DNA polymerase slips or stutters and the newly synthesized strand shifts on the template and reanneals to an adjacent repetitive element, producing unpaired loop intermediates localized to one or the other strands. If this unpaired loop intermediate is not repaired then it will be perpetuated as an insertion of an instance of the repetitive element if in the daughter strand, or a deletion if in the template strand.

*A Possible Correlation to Intrinsic Hotspots.* A higher frequency of somatic hypermutation has been reported to occur at sequence motifs referred to as intrinsic hotspots (for review see reference 12). Interestingly, every insertion/deletion event reported in our study resembled one of these hotspots (AGC, TAC, and AAC; references 12 and 27; Fig. 4). The analysis of selected populations may have influenced this tendency because seven out of eight of these events occurred in the CDRs where it has been shown that hotspot motifs are preferentially found (25, 26). Furthermore, only a weak correlation to hotspots could be found for the previously reported insertions/deletions involving unselected regions of V loci (Table 3). However, the single event found in this analysis that occurred outside of the CDRs in FW3 (clone tm121, Figs. 1 *C* and 4 *A*), also involved a tandem of possible hotspots (AAG, AAC). A more extensive and directed analysis is required to fully address this issue.

*Implications for the Molecular Mechanism of Somatic Hypermutation.* The instability of repetitive tracts during DNA replication is a hallmark of defects in postreplicative mismatch repair (33), and the locus-specific downregulation of DNA mismatch repair in response to UV irradiation has recently been reported for immunoglobulin $V_H$ genes in freshly sorted GC B cells ($CD38^+IgD^-$) compared to mantle zone B cells ($CD38^-IgD^+$; reference 34). In a recent study by Tran et al. (35), it was shown that tract instability of homonucleotide runs associated with mismatch repair defects occur more frequently in long than in short runs. These authors suggested that if loop intermediates occur in long repetitive tracts (>8 bp for a homonucleotide run) they could involve a distal repetitive element out of reach of the polymerase proofreading activity and only be subjected to mismatch repair. However, for short repetitive tracts, as for the events reported in this analysis, loop intermediates can only occur proximal to the polymerase complex and are therefore subjected to both polymerase proofreading and mismatch repair mechanisms.

All 9 events in this analysis, and 19 out of 25 events from the literature (28 out of 34 insertions and deletions reported), appeared to result from secondary structural intermediates. Loop intermediates proximal to the polymerase complex during DNA polymerization should be repaired by the polymerase proofreading mechanisms immediately, or by the postreplicative DNA repair systems. This analysis suggests the following characteristics for the polymerization process during somatic hypermutation. (*a*) The polymerase interacts with the V locus in a particularly unstable or "loose" fashion, especially when hotspot motifs or elements capable of forming secondary structures are encountered, allowing bp substitutions in most instances, and insertions or deletions via polymerase slippage at a much lower frequency; (*b*) it has limited proofreading capabilities; and (*c*) there is a downregulation of postreplicative mismatch repair. An efficient means to downregulate mismatch repair during somatic hypermutation could be through the lack of differentiation of the template and progeny strands for the mismatch repair system; lack of strand differentiation has been shown to increase the rate of mutations introduced (36). Such a system would be advantageous for the locus-specific V gene somatic hypermutation in that it could involve alterations of a single enzymatic complex (polymerase complex) rather than multiple systems (proofreading and mismatch repair). Another system, which would have the

same advantage, i.e., the alteration of a single complex, would be the alteration of a DNA repair system such as transcription-coupled repair to be the somatic mutator, as suggested in recent studies (13). Alternatively, the insertions and deletions might result solely from a downregulation of postreplicative mismatch repair at the V locus in the rapidly proliferating centroblasts that are undergoing somatic hypermutation or due to a polymerase enzyme with such a high fault rate as to overwhelm any repair.

All currently accepted models of somatic hypermutation, whether related to DNA excision-repair–like systems or transcription-repair, or to DNA polymerization or reverse transcription, involve transcriptional activation involving *cis*-factors in the V locus (enhancers, etc.) followed by the activity of unknown polymerase enzymes of some type. This analysis does not refute or corroborate any of these models directly, but it does provide further characterization of the polymerization system involved, based on the types of mutations observed and on the molecular biology that is known to cause such mutations. This analysis and the model presented here provide further information or criteria to be contemplated as the various possible polymerase systems involved are considered.

*Conclusions.* Insertions and deletions into immunoglobulin $V_H$ genes during somatic hypermutation are additional means by which the immunoglobulin repertoire can be diversified. These events display characteristics supporting models of somatic hypermutation involving a particularly unstable or error-prone polymerase to allow the introduction of mutations, and involving the downregulation of DNA repair to allow the perpetuation of these mutations. Additionally, we show that these events tend to involve sequence motifs resembling intrinsic hotspots of somatic hypermutation, suggesting that the polymerase complex is destabilized in a sequence-specific manner to allow preferential mutation at these sequence elements.

We are grateful to Yucheng Li, Fang Zhao, Steve Scholl, Carol Williams, Shirley Hall, and Robin Wray, all of whom provided unprecedented technical assistance for various aspects of this work. We also thank Kimble Frazer for excellent discussions.

These studies were supported by a grant from the National Institutes of Health (AI-12127). J. Donald Capra holds the Edwin L. Cox Distinguished Chair in Immunology and Genetics.

Address correspondence to Dr. Virginia Pascual, Molecular Immunology Center, Department of Microbiology, UT Southwestern Medical Center, 6000 Harry Hines Blvd., Dallas, Texas 75235-9140. Phone: 214-648-1918; Fax: 214-648-1915; E-mail: vpascu@mednet.swmed.edu

*Received for publication 14 July 1997 and in revised form 1 October 1997.*

## References

1. Clark, E.A., and J.A. Ledbetter. 1994. How B and T cells talk to each other. *Nature.* 367:425–428.
2. Dorken, B., P. Moller, A. Pezzutto, R. Schwartz-Albiez, and C. Moldenhauer. 1989. B-cell antigens. *In* Leukocyte Typing IV. A.J. McMichael, editor. Oxford University Press, London. pp. 131–140.
3. Ling, N.R., I.C.M. MacLennan, and D.Y. Mason. 1987. B-cell and plasma cell antigens: new and previously defined clusters. *In* Leukocyte Typing III. A.J. McMichael, editor. Oxford University Press, London. pp. 302–308.
4. Pascual, V., Y.J. Liu, A. Magalski, O. de Bouteiller, J. Banchereau, and J.D. Capra. 1994. Analysis of somatic mutation in five B cell subsets of human tonsil. *J. Exp. Med.* 180: 329–339.
5. Liu, Y.J., O. de Bouteiller, C. Arpin, F. Briere, L. Galibert, S. Ho, H. Martinez-Valdez, J. Banchereau, and S. Lebecque. 1996. Normal human $IgD^+IgM^-$ germinal center B cells can express up to 80 mutations in the variable region of their IgD transcripts. *Immunity.* 4:603–613.
6. Denepoux, S., D. Razanajaona, D. Blanchard, G. Meffre, J.D. Capra, J. Banchereau, and S. Lebecque. 1997. Induction of somatic mutation in a human B cell line in vitro. *Immunity.* 6:35–46.
7. Berek, C. 1993. Somatic mutation and memory. *Curr. Opin. Immunol.* 5:218–222.
8. Kelsoe, G. 1996. Life and death in germinal centers (redux). *Immunity.* 4:107–111
9. Liu, Y.J., G.D. Johnson, J. Gordon, and I.C.M. MacLennan. 1992. Germinal centres in T-cell-dependent antibody responses. *Immunol. Today.* 13:17–21.
10. MacLennan, I.C.M., Y.J. Liu, and G.D. Johnson. 1992. Maturation and dispersal of B-cell clones during T cell–dependent antibody responses. *Immunol. Rev.* 126:143–161.
11. McKean, D., K. Huppi, M. Bell, L. Staudt, W. Gerhard, and M. Weigert. 1984. Generation of antibody diversity in the immune response of BALB/c mice to influenza virus hemagglutinin. *Proc. Natl. Acad. Sci. USA.* 81:3180–3186.
12. Neuberger, M.S., and C. Milstein. 1995. Somatic hypermutation. *Curr. Opin. Immunol.* 7:248–254.
13. Storb, U. 1996. The molecular basis of somatic hypermutation of immunoglobulin genes. *Curr. Opin. Immunol.* 8:206–214.
14. Liu, Y.J., O. de Bouteiller, C. Arpin, I. Durand, and J. Banchereau. 1994. Five human mature B cell subsets. *Adv. Exp. Med. Biol.* 355:289–296.
15. Marks, J.D., M. Tristem, A. Karpas, and G. Winter. 1991. Oligonucleotide primers for polymerase chain reaction amplification of human immunoglobulin variable genes and design of family-specific oligonucleotide probes. *Eur. J. Immunol.* 21:985–991.
16. Sanz, I., P. Kelly, C. Williams, S. Scholl, P. Tucker, and J.D. Capra. 1989. The smaller human $V_H$ gene families display remarkably little polymorphism. *EMBO (Eur. Mol. Biol. Organ.) J.* 8:3741–3448.
17. Potter, K.N., Y.C. Li, and J.D. Capra. 1994. The cross-reac-

tive idiotopes recognized by the monoclonal antibodies 9G4 and LC1 are located in framework region 1 of two non-overlapping subsets of human $V_H4$ family encoded antibodies. *Scand. J. Immunol.* 40:43–49.

18. Tomlinson, I.M., G. Walter, J.D. Marks, M.B. Llewelyn, and G. Winter. 1992. The repertoire of human germline $V_H$ sequences reveals about fifty groups of $V_H$ segments with different hypervariable loops. *J. Mol. Biol.* 227:776–798.
19. Umar, A., and P.J. Gearhart. 1995. Reciprocal homologous recombination in or near antibody VDJ genes. *Eur. J. Immunol.* 25:2392–2400.
20. Ford, J.E., M.G. McHeyzer-Williams, and M.R. Lieber. 1994. Chimeric molecules created by gene amplification interfere with the analysis of somatic hypermutation of murine immunoglobulin genes. *Gene.* 142:279–283.
21. Andris, J.S., B.R. Brodeur, and J.D. Capra. 1993. Molecular characterization of human antibodies to bacterial antigens: utilization of the less frequently expressed $V_H2$ and $V_H6$ heavy chain variable region gene families. *Mol. Immunol.* 30: 1601–1616.
22. Brodeur, B.R., J. Hamel, D. Martin, and P. Rondeau. 1991. Biological activity of a human monoclonal antibody to *Bordetella pertussis* lipooligosaccharide. *Hum. Antib. Hybrid.* 2:194–199.
23. Betz, A.G., C. Rada, R. Pannell, C. Milstein, and M.S. Neuberger. 1993. Passenger transgenes reveal intrinsic specificity of the antibody hypermutation mechanism: clustering, polarity, and specific hot spots. *Proc. Natl. Acad. Sci. USA.* 90: 2385–2388.
24. Rogozin, I.B., and N.A. Kolchanov. 1992. Somatic hypermutagenesis in immunoglobulin genes. II. Influence of neighbouring base sequences on mutagenesis. *Biochim. Biophys. Acta.* 1171:11–18.
25. Goyenechea, B., and C. Milstein. 1996. Modifying the sequence of an immunoglobulin V-gene alters the resulting pattern of hypermutation. *Proc. Natl. Acad. Sci. USA.* 93:13979–13984.
26. Wagner, S.D., C. Milstein, and M.S. Neuberger. 1995. Codon bias targets mutation. *Nature.* 376:732–733.
27. Smith, D.S., G. Creadon, P.K. Jena, J.P. Portanova, B.L. Kotzin, and L.J. Wysocki. 1996. Di- and trinucleotide target preferences of somatic mutagenesis in normal and autoreactive B cells. *J. Immunol.* 156:2642–2652.
28. Potter, K.N., Y. Li, and J.D. Capra. 1996. Staphylococcal protein A simultaneously interacts with framework region 1, complementarity-determining region 2, and framework region 3 on human $V_H3$-encoded Igs. *J. Immunol.* 157:2982–2988.
29. Golding, G.B., P.J. Gearhart, and B.W. Glickman. 1987. Patterns of somatic mutations in immunoglobulin variable genes. *Genetics.* 115:169–176.
30. Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. This paper is dedicated to Professor Theodosius Dobzhansky on the occasion of his 66th birthday. *Cold Spring Harbor. Symp. Quant. Biol.* 31:77–84.
31. Ripley, L.S. 1990. Frameshift mutation: determinants of specificity. *Annu. Rev. Genet.* 24:189–213.
32. Modrich, P., and R. Lahue. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65:101–133.
33. Strand, M., T.A. Prolla, R.M. Liskay, and T.D. Petes. 1993. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature.* 365: 274–276.
34. Fairhurst, R.M., Y. Valles-Ayoub, M. Neshat, and J. Braun. 1996. A DNA repair abnormality specific for rearranged immunoglobulin variable genes in germinal center B cells. *Mol. Immunol.* 33:231–244.
35. Tran, H.T., J.D. Keen, M. Kricker, M.A. Resnick, and D.A. Gordenin. 1997. Hypermutability of homonucleotide runs in mismatch repair and DNA polymerase proofreading yeast mutants. *Mol. Cell Biol.* 17:2859–2865.
36. MacPhee, D.G. 1996. Mismatch repair as a source of mutations in non-dividing cells. *Genetica. (The Hague).* 97:183–195.
37. Matsuda, F., and T. Honjo. 1996. Organization of the human immunoglobulin heavy-chain locus. *Adv. Immunol.* 62:1–29.
38. Dunn-Walters, D.K., P.G. Isaacson, and J. Spencer. 1997. Sequence analysis of human $IgV_H$ genes indicates that ileal lamina propria plasma cells are derived from Peyer's patches. *Eur. J. Immunol.* 27:463–467.
39. Both, G.W., L. Taylor, J.W. Pollard, and E.J. Steele. 1990. Distribution of mutations around rearranged heavy-chain antibody variable-region genes. *Mol. Cell Biol.* 10:5187–5196.
40. Allen, D., T. Simon, F. Sablitzky, K. Rajewsky, and A. Cumano. 1988. Antibody engineering for the analysis of affinity maturation of an anti-hapten. *EMBO (Eur. Mol. Biol. Organ.) J.* 7:1995–2001.
41. Kim, S., M. Davis, E. Sinn, P. Patten, and L. Hood. 1981. Antibody diversity: somatic hypermutation of rearranged $V_H$ genes. *Cell.* 27:573–581.
42. Lebecque, S.G., and P.J. Gearhart. 1990. Boundaries of somatic mutation in rearranged immunoglobulin genes: 5′ boundary is near the promoter, and 3′ boundary is approximately 1 kb from V(D)J gene. *J. Exp. Med.* 172:1717–1727.
43. Kataoka, T., T. Nikaido, T. Miyata, K. Moriwaki, and T. Honjo. 1982. The nucleotide sequences of rearranged and germline immunoglobulin $V_H$ genes of a mouse myeloma MC101 and evolution of $V_H$ genes in mouse. *J. Biol. Chem.* 257:277–285.
44. Gearhart, P.J., and D.F. Bogenhagen. 1983. Clusters of point mutations are found exclusively around rearranged antibody variable genes. *Proc. Natl. Acad. Sci. USA.* 80:3439–3443.
45. Rickert, R. and S. Clarke. 1993. Low frequencies of somatic mutation in two expressed V kappa genes: unequal distribution of mutation in 5′ and 3′ flanking regions. *Int. Immunol.* 5:255–263.
46. Weber, J.S., J. Berry, S. Litwin, and J.L. Claflin. 1991. Somatic hypermutation of the JC intron is markedly reduced in unrearranged kappa and H alleles and is unevenly distributed in rearranged alleles. *J. Immunol.* 146:3218–3226.
47. Rogerson, B., J. Hackett, Jr., A. Peters, D. Haasch, and U. Storb. 1991. Mutation pattern of immunoglobulin transgenes is compatible with a model of somatic hypermutation in which targeting of the mutator is linked to the direction of DNA replication. *EMBO (Eur. Mol. Biol. Organ.) J.* 10:4331–4341.
48. Kosmas, C., N.A. Viniou, K. Stamatopoulos, N.S. Courtenay-Luck, T. Papadaki, P. Kollia, G. Paterakis, D. Anagnostou, X. Yataganas, and D. Loukopoulos. 1996. Analysis of the kappa light chain variable region in multiple myeloma. *Br. J. Haematol.* 94:306–317.
49. Wu, H.Y., and M. Kaartinen. 1995. The somatic hypermutation activity of a follicular lymphoma links to large insertions and deletions of immunoglobulin genes. *Scand. J. Immunol.* 42:52–59.